

This disclosure describes the classes of tools that permit a scientist to generate experimentally testable hypotheses concerning the function of a protein starting from an evolutionary analysis. These are outlined below:

I. Tools that detect change in function within a family of proteins.

- A. Ratios of silent to non-silent substitution along specific branches of an evolutionary tree including tools that address normalization issues.
- B. Covarion behavior, in which individual residues display different mutability in different branches of a tree.
- C. Detecting high absolute rates of amino acid substitution, changes per unit time.
- II. Tools that detect conservation of function within a family of proteins.
  - A. Compensatory changes
  - B. Homoplasy
  - C. Absolute conservation within a defined evolutionary distance
- III. Tools that identify individual residues involved in changes in functionally significant behavior.
  - A. Residues changing in episodes with high Ka/Ks values, minus residues changing in episodes with low Ka/Ks values
  - B. Residues displaying covarion behavior
  - C. Mapping these residues on to models for the secondary, tertiary, and quaternary structure of proteins.
- IV. Tools that identify individual residues involved in conserved of functionally significant behavior
  - A. Residues suffering compensatory changes
  - B. Residues displaying homoplasy
  - C. Mapping these residues on to models for the secondary, tertiary, and quaternary structure of proteins.
- V. Tools that involve correlation between the evolutionary histories of two families of proteins
  - A. Correlating the topology of evolutionary trees in two families of proteins
  - B. Correlating the connectivity of proteins in a gene family
  - C. Dating events in the molecular history
  - D. Correlating evolutionary events in two protein families occuring at approximately the same time
  - E. Correlating evolutionary events in two protein families that are associated with analogous behavior involving expressed/silent ratios
- VI. Tools that involve correlation between the evolutionary history of a family of proteins and the evolutionary history of the organism as known from some source other than genomic sequence data, including paleontology, geology, ecology, ontogeny, phylogeny, or systematics (collectively known as the "non-genomic record".
  - A. Correlating the topology of an evolutionary trees and the non-genomic record.
  - B. Correlating features of patterns of evolution in specific branches in the evolutionary tree with the non-genomic record
  - C. Correlating evolutionary events in several protein families occurring at approximately the same time with the non-genomic record